Evaluating the credibility of transport processes in simulations of ozone recovery using the Global Modeling Initiative three-dimensional model

Susan E. Strahan
Goddard Earth Science and Technology Center, University of Maryland Baltimore County, Baltimore, Maryland, USA

Anne R. Douglass
NASA Goddard Space Flight Center, Greenbelt, Maryland, USA

Received 10 October 2003; revised 24 December 2003; accepted 22 January 2004; published 13 March 2004.

[1] The Global Modeling Initiative (GMI) has integrated two 36-year simulations of an ozone recovery scenario with an offline chemistry and transport model using two different meteorological inputs. Physically based diagnostics, derived from satellite and aircraft data sets, are described and then used to evaluate the realism of temperature and transport processes in the simulations. Processes evaluated include barrier formation in the subtropics and polar regions, and extratropical wave-driven transport. Some diagnostics are especially relevant to simulation of lower stratospheric ozone, but most are applicable to any stratospheric simulation. The global temperature evaluation, which is relevant to gas phase chemical reactions, showed that both sets of meteorological fields have near climatological values at all latitudes and seasons at 30 hPa and below. Both simulations showed weakness in upper stratospheric wave driving. The simulation using input from a general circulation model (GMI_{GCM}) showed a very good residual circulation in the tropics and Northern Hemisphere. The simulation with input from a data assimilation system (GMI_{DAS}) performed better in the midlatitudes than it did at high latitudes. Neither simulation forms a realistic barrier at the vortex edge, leading to uncertainty in the fate of ozone-depleted vortex air. Overall, tracer transport in the offline GMI_{GCM} has greater fidelity throughout the stratosphere than it does in the GMI_{DAS}. *INDEX TERMS:* 0341 Atmospheric Composition and Structure: Middle atmosphere—constituent transport and chemistry (3334); 3319 Meteorology and Atmospheric Dynamics: General circulation; 3337 Meteorology and Atmospheric Dynamics: Numerical modeling and data assimilation; *KEYWORDS:* stratospheric model, process evaluation, transport

Citation: Strahan, S. E., and A. R. Douglass (2004), Evaluating the credibility of transport processes in simulations of ozone recovery using the Global Modeling Initiative three-dimensional model, *J. Geophys. Res.*, *109*, D05110, doi:10.1029/2003JD004238.

1. Introduction

[2] For the past few decades, chemistry and transport models have been used to assess the impact of natural and anthropogenic perturbations such as aircraft emissions or chlorofluorocarbon growth on stratospheric ozone. Most assessments relied on two-dimensional (zonally averaged) models that cannot physically represent inherently three-dimensional (3-D) processes such as transport out of the polar vortex and cross tropopause transport [*Park et al.*, 1999]. Some recent efforts have used 3-D chemistry and transport models (CTMs), which provide more realistic representations of nonzonal processes [*Danilin et al.*, 1998; *Douglass et al.*, 1999; *Kinnison et al.*, 2001], although the third dimension greatly increases the computational requirements and demands greater human resources for evaluation. In spite of this, development of a 3-D assessment model is a worthwhile goal because it offers the opportunity to improve the physical basis of assessment modeling and, if the 3-D model compares well against observations, reduce uncertainties due to transport.

[3] The Global Modeling Initiative (GMI) was formed in 1995 with the goal of producing a well-tested 3-D chemistry and transport model that could be used for assessments and other controlled experiments that required a common framework. In their first experiments, hereinafter referred to as GMI-1, the GMI science team tested advection schemes, chemical mechanisms, solvers, and meteorological inputs to determine the CTM modules that would produce the most realistic assessment model [*Rotman et al.*, 2001]. To establish which simulation was the most credible, physically based tests, derived from aircraft and satellite data sets, were used to evaluate simulations using different meteorological inputs [*Douglass et al.*, 1999]. Six tests were created, each evaluating a different aspect of stratospheric transport and

0148-0227/04/2003JD004238

mixing, and grading standards were defined by the observations and their uncertainties. The simulations received scores on each test that could then be used to quantitatively distinguish between them. At the end of the evaluation, the GMI science team could objectively select the best meteorological data set for simulating the effects of supersonic aircraft on the stratosphere. The modules selected were then used to simulate an aircraft emissions scenario, described by *Kinnison et al.* [2001].

[4] Recently, the GMI science team ran two 36-year integrations of the World Meteorological Organization (WMO) scenario "MA2" [*World Meteorological Organization* (*WMO*), 2002] with the GMI-CTM. During the period simulated, 1995–2030, the scenario's organic chlorine and bromine boundary conditions decline while N_2O and CH_4 increase. While the intent of this WMO scenario is to predict future ozone change, the primary purpose of this GMI study is to assess the sensitivity of model predictions to differences in transport. Transport in a CTM depends on meteorological input; for this scenario we chose meteorological fiends from the Finite Volume General Circulation Model (FVGCM) and from the Finite Volume Data Assimilation System (FVDAS). These models were selected because although they have significant differences in residual circulation and mixing [*Schoeberl et al.*, 2003], each is able to realistically represent some aspects of the stratosphere. Our choice of the particular model year was based on that year having a cold Arctic lower stratospheric winter compared to actual climatology because we wished to examine transport sensitivities in a "worst case scenario" for ozone recovery. Each 36-year GMI-CTM simulation was produced using the same year of meteorological input (from the FVGCM or the FVDAS) for all 36 years while the trace gas boundary conditions changed. These CTM simulations will be referred to as GMI_{GCM} and GMI_{DAS}. Constituent transport with these meteorological fields was initially evaluated using the Goddard CTM, a model with the same advection core as the GMI-CTM. This evaluation used the GMI-1 grading criteria described by *Douglass et al.* [1999] and, in comparison with the test results presented in that paper, showed that the stratospheric transport characteristics of the FVGCM and FVDAS meteorological fields were superior to meteorological fields used in the GMI-1 CTM simulations.

[5] The goal of this paper is to present observationally based diagnostics for the evaluation of specific stratospheric processes and features; philosophically, this paper builds on the diagnostic work of *Douglass et al.* [1999] and *Rotman et al.* [2001]. The emphasis here is not just "comparison with data," but (1) identification of atmospheric processes relevant to the realistic simulation of a phenomenon or feature and (2) identification of a data set that demonstrates the occurrence of this process. Such a data set becomes the physical basis for model evaluation. A model earns credibility when it can be shown to realistically represent known atmospheric processes. Some of the atmospheric processes relevant to the WMO MA2 scenario can be tested using previously defined GMI evaluation criteria. In this paper, new physically based stratospheric diagnostics are presented that illustrate additional stratospheric processes. While some of the tests are especially relevant to this scenario with declining halogens, all tests are generally applicable to any stratospheric simulation. In the following sections, we identify some important stratospheric processes and their diagnostic tests, then apply them to two GMI 3-D-CTM simulations in order to evaluate many (though not all) aspects affecting their credibility in an ozone recovery scenario. Applying these tests to simulations run in a common framework (i.e., the GMI-CTM) allows us to examine the sensitivity of the results to the meteorological input.

2. Evaluating the Suitability of GMI Simulations for Use in Ozone Predictions

[6] How do you evaluate a model's credibility? The GCM-Reality Intercomparison Project for SPARC (GRIPS, where SPARC is Stratospheric Processes And their Role in Climate) compared temperature and wind fields in 13 3-D middle atmosphere climate models against observations to identify deficiencies in dominant atmospheric features, such as the location of the jets and polar temperatures [*Pawson et al.*, 2000]. To evaluate an assessment model, one might look at qualitative agreement with historical ozone trends. The Scientific Assessment of Ozone Depletion [*WMO*, 2002] shows many simulations of column ozone. While many models show qualitative agreement with historical ozone from 1980 to 2000, their predictions of future ozone diverge. Such agreement is a misleading diagnostic since the total column represents the integrated effects of chemistry and transport at many altitudes. To understand the difference in performance between eight chemistry-climate models, *Austin et al.* [2003] chose several specific diagnostics, such as ozone climatology, polar temperature biases, and poleward heat fluxes. Their intent, commensurate with that of GMI, is to identify diagnostics for processes influencing ozone and use them to reduce the uncertainty in predictions of future ozone levels.

[7] In this study we assess model credibility by evaluating temperature and transport processes. This is done with observationally based tests at a variety of altitudes and latitudes. In GMI-1, we developed tests to assess model transport processes in regions that would be perturbed by stratospheric aircraft exhaust. Since emissions were projected to occur in the upper troposphere and lower stratosphere, transport near the tropopause was especially important and tests were developed that emphasized model fidelity there. In this scenario with decreasing organic halogens and increasing N_2O, transport fidelity is important at all levels in the stratosphere; Cl changes affect O_3 losses in the upper and lower stratosphere, while the changing NO_x abundance will have an impact on O_3 loss in the middle stratosphere. In this section, new tests are presented that expand the scope of the GMI's stratospheric evaluation. These tests, as well as some of the previous GMI-1 tests, are applied to two new GMI-CTM simulations.

2.1. Models and Data Sets

[8] The GMI chemistry and transport model uses the flux form semi-Lagrangian transport scheme of *Lin and Rood* [1996]. The simulations were run at a resolution of 4° latitude by 5° longitude with 28 vertical levels (model lid at 0.43 hPa) using a sigma-pressure coordinate. The input FVGCM and FVDAS winds field both have an original

resolution of 2° latitude by 2.5° longitude and 55 vertical levels (to 0.01 hPa); the horizontal resolution was degraded and the original vertical levels were mapped to 29 levels using a divergence conserving algorithm in order to make a 36-year simulation tractable [*Rotman et al.*, 2001]. Meteorological input is updated every six hours. This CTM uses the SMVGEAR II chemical solver and is an improved version of that described by *Douglass et al.* [1999] and *Rotman et al.* [2001]. Individual species are advected separately, with the exception of atomic radicals. The parameterization for the effects of polar stratospheric clouds is described by *Considine et al.* [2000], and the effects of polar processes in these two simulations are discussed in detail by D. B. Considine et al. (Sensitivity of Antarctic ozone recovery predictions to GCM and DAS dynamics, submitted to *Journal of Geophysical Research*, 2004) (hereinafter referred to as Considine et al., submitted manuscript, 2004). Additional details on the GMI model and input meteorological fields are described by A. R. Douglass et al. (Radicals and reservoirs in the Global Modeling Initiative chemistry and transport model: Comparison to measurements, submitted to *Journal of Geophysical Research*, 2004) (hereinafter referred to as Douglass et al., submitted manuscript, 2004).

[9] Data sets used in these tests are from the National Center for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) Reanalysis, the Upper Atmosphere Research Satellite (UARS), and an airborne spectrometer. NCEP/NCAR temperature reanalyses from 1980 to 1999 are used to create a climatology of monthly temperature distributions for 8 stratospheric levels and 11 latitude bands. (See *Newman et al.* [2001] for details of the reanalysis products.) UARS data sets include N_2O and CH_4 from the Cryogenic Limb Array Etalon Spectrometer (CLAES) [*Roche et al.*, 1996] and CH_4 from the Halogen Occultation Experiment (HALOE) [*Park et al.*, 1996]. Both instruments began operation October 1991; CLAES made measurements for about 18 months and HALOE has operated nearly continuously since UARS was launched. CLAES has high-spatial-density sampling that alternates between 35°S–80°N and 35°N–80°S every 35 days (the yaw maneuver); CH_4 and N_2O retrievals have a vertical resolution of ~3 km and offer their best precision at 46 hPa and above [*Roche et al.*, 1996]. HALOE collects about 30 profiles daily with latitudes ranging from 80°S–80°N; latitudes poleward of 60° are only sampled in spring and summer. HALOE CH_4 data have a vertical resolution of 4 km and maintain their precision (7%) throughout the stratosphere, down to 100 hPa [*Park et al.*, 1996]. HALOE data are especially useful for investigating interannual variability and lower stratospheric transport.

2.2. Generalized Tests of Stratospheric Temperature and Transport

[10] This section examines basic features and processes such as temperature, transport, and mixing. Section 2.3 presents tests that are relevant to processes affecting polar ozone.

2.2.1. Global Temperature

[11] While both gas and heterogeneous phase reactions depend on temperature, different tests are required to evaluate how model temperatures affect them. Gas phase temperature-dependent reactions will proceed at a slightly slower or faster rate as temperature varies, but heterogeneous reactions only occur if the necessary temperature threshold is reached. Here, we evaluate model temperatures with regard to gas phase chemical reactions by comparing how often and how closely they agree with climatological values. This is accomplished by comparing the model and observed most probable temperatures for a given month, latitude range, and altitude. Other evaluations of climate models have focused on polar temperatures or have looked at broadly averaged (i.e., monthly, annual, or global) temperatures [*Austin et al.*, 2003; *Pawson et al.*, 2000]. The temperature diagnostic shown here attempts to make a spatially and temporally thorough comparison using as little averaging as possible, whose results can be displayed as simply as possible. Model temperatures in the GMI simulations are a property of the input meteorological fields (i.e., of the FVGCM and the FVDAS).

[12] Stratospheric temperature evaluation is based on the distribution of NCEP/NCAR reanalysis temperatures from 1980 to 1999. Daily area-weighted temperature distributions are first calculated for 11 latitude bands and 8 pressure levels from 150 to 1 hPa for the entire 20-year data set, and then all distributions from each month are combined to create a 20-year climatology of monthly temperature distributions. The test itself examines the difference between the model and climatological most probable temperature (MPT) for each month and latitude band, resulting in 132 points of comparison on each of 8 pressure surfaces. Figure 1 shows how the FVGCM and FVDAS MPTs differ from the NCEP/NCAR values on the 50 hPa and 5 hPa surfaces. At 50 hPa, both simulations do an excellent job of producing climatological temperatures; most differences range from 0 to 3 K. At 5 hPa, both simulations are too cool, but the FVDAS is generally about 3 K closer to the climatological temperatures than the

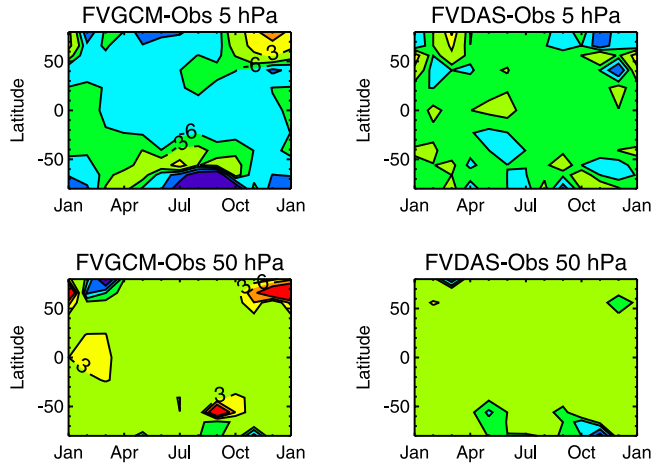

Figure 1. Difference between the models' and the 20-year NCEP climatological most probable temperatures on two surfaces. Most probable values are calculated monthly for 11 latitude bands. Contour intervals are 3 K. At 50 hPa both simulations are usually within 0–3 K of the climatological value. The top panels show a cold bias in the upper stratosphere in both simulations. The FVDAS bias is typically about 3 K less than the FVGCM.

FVGCM. Also, there is no apparent pattern to the FVDAS differences while the FVGCM's worst agreement proceeds from northern spring, through the tropics, to southern spring.

[13] It is important to remember that this is not a comparison of model climatology with NCEP climatology. FVDAS temperatures were assimilated for the period July 1999 to June 2000 while the FVGCM temperatures represent one year of a 36-year GCM simulation. This test is a general reality check to ensure that the model years chosen are within observed climatology. The test also enables us to identify regions of model bias.

[14] To summarize how each model level compares with observations, an area-weighted distribution of the differences between one year of model and climatological most probable temperatures are plotted in Figure 2. The FVDAS consistently produces MPTs in better agreement with climatology than the FVGCM. For the 8 pressure levels tested, the FVDAS is within 3 K of climatology 82% of the time, while the FVGCM agrees within 3 K 69% of the time. The results at 100 hPa are quite interesting because this is the only level with a bimodal distribution of differences. A contour plot of the MPT differences as a function of latitude and time (not shown) shows that both simulations have a bias toward low temperatures near the tropical tropopause, while poleward of 30° each model has excellent agreement with climatology.

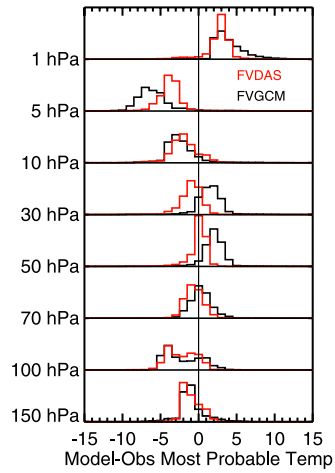

Figure 2. Summary of the models' temperature behavior with respect to the 20-year NCEP climatology. Each histogram gives the area-weighted difference between model and climatological most probable temperatures (MPTs) on a given pressure surface for all latitude bands and months of the year. (Each histogram contains 132 MPT comparisons.) Overall, the FVDAS has temperatures closer to the 20-year climatological means than does the FVGCM.

2.2.2. Residual Circulation

[15] In the following subsections we evaluate the effects of the residual circulation on constituent distributions in the tropics and extratropics. The quasi-biennial oscillation (QBO) is known to create a secondary circulation that affects trace gas distributions in the tropics and midlatitudes [*Gray and Russell*, 1999]. Most GCMs cannot produce a QBO and there is no value in creating a test that nearly all models will fail. Transport tests should either carried out at latitudes poleward of the effects of the QBO or they should diagnose the time-averaged behavior of transport in QBO-influenced regions. To this end, the tests in section 2.2.2.1 use data poleward of 44° where interannual variability is low; in section 2.2.2.2, the test includes profile data from both QBO phases; and in section 2.2.2.3, the feature diagnosed does not depend on the phase of the QBO or on the latitude of the subtropical boundary, which varies with the phase of the QBO.

2.2.2.1. Annual Cycle of CH_4 in the Extratropical Middle and Upper Stratosphere

[16] Transport has an important effect on the ozone distribution of the middle and upper stratosphere even though photochemical timescales are relatively short there. For example, NO_x family chemistry dominates O_3 loss in the sunlit middle stratosphere, but NO_x mixing ratios depend strongly on NO_y abundance, which is largely controlled by transport. The dynamics and composition of the upper stratosphere also affect ozone in the polar lower stratosphere because the strength of wave activity aloft determines descent rates and influences the fraction of upper stratospheric air reaching the lower stratosphere by the end of winter [*Rosenfield and Schoeberl*, 2001].

[17] We evaluate model transport and mixing processes affecting the upper stratosphere through analysis of the extratropical CH_4 annual cycle. In the high latitudes it is a function of seasonally varying meridional transport, mixing, and descent. Photochemistry also matters in summer. The amplitude, phase, and variability of CH_4 annual cycle provide useful measures of the timing and strength of transport and the presence of photochemical processes. Previous work by *Randel et al.* [1998] combined HALOE and CLAES CH_4 measurements to estimate the annual cycle at high latitudes, which they interpreted in terms of seasonal variations in dynamical and photochemical forcing. HALOE CH_4 data show large interannual variability equatorward of 44° due to QBO influence [*Randel et al.*, 1998], so we choose to study only the middle- and high-latitude ranges, 44°–56° and 72°–80°, which show much lower interannual variability. CLAES CH_4 is used because HALOE does not sample the polar region in fall and winter.

[18] This test compares several features of the CH_4 annual cycle at 800 K (~10 hPa) and 1200 K (~5 hPa). Probability distribution functions (pdfs) calculated from daily 1992 CLAES data within each latitude band are contoured together to show the amplitude, phase, and variability of the cycle. Figure 3 shows a 1-year cycle of contoured pdfs from CLAES, the GMI_{GCM}, and the GMI_{DAS} at 1200 K for 4 extratropical latitude bands (two in each hemisphere). The 35-days gaps in the data are caused by UARS yaw maneuvers. A narrow distribution means a high probability of a narrow range of mixing ratios (yellow and red contours) and indicates a homogeneous atmosphere; this may result from rapid photochemistry or mixing. A broad distribution (purple and blue contours) indicates that transport from a photochemically different region dominates processes that homogenize the atmosphere. The four diagnostic quantities derived from CH_4 annual cycles and the grading of model results are described below; the GMI_{GCM} and GMI_{DAS} scores are listed in Table 1.

[19] The first diagnostic quantity is the most probable CH_4 mixing ratios. We compare the most probable value

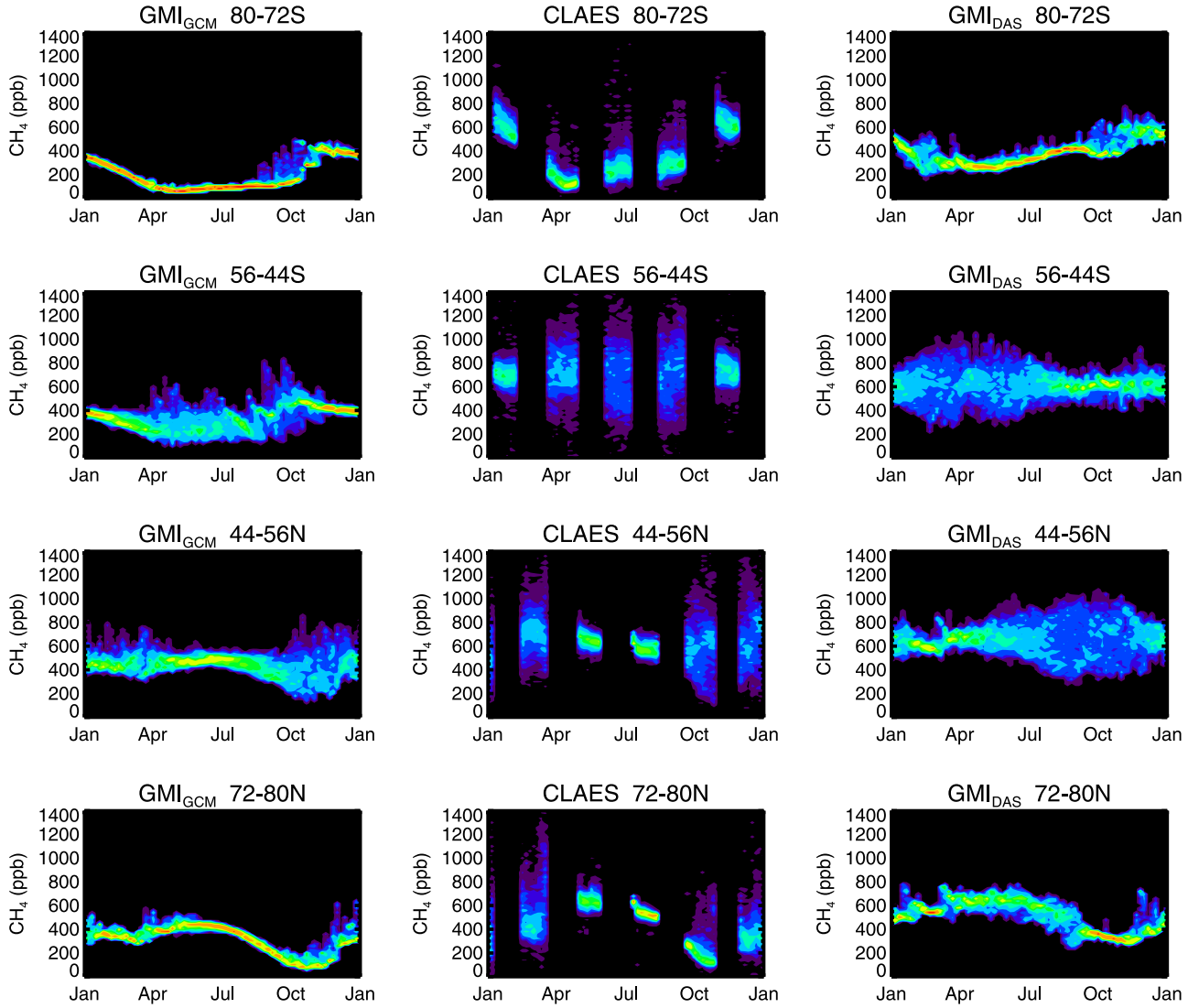

Figure 3. Comparison of CLAES CH_4 extratropical annual cycles with GMI_{GCM} and GMI_{DAS} for four latitude bands on the 1200 K surface. The CLAES cycles are produced from contours of daily CH_4 pdfs for all available dates in 1992; the models' pdfs are calculated every third day. Yellow and red indicate frequent occurrence of that mixing ratio, implying a well-mixed distribution. Blue and purple, which represent low probability, are usually part of broad distributions. Broad distributions arise when long-range transport dominates processes that reduce variability, namely, rapid photochemistry and mixing.

rather than the mean in order to assess a region's overall composition in a way that is not affected by the spatial distribution of the tracer [*Sparling*, 2000]. Because tracer mixing ratios depend on various transport processes as well as on chemistry, this test assesses the general balance between model transport and chemistry. In a model, particularly in the upper stratosphere, tracer mixing ratios may be affected by the height of the model lid. For example, if the lid is near the stratopause, CH_4 in the winter polar stratosphere may be too high because the model lacks a mesospheric source of low CH_4 (D. Waugh, personal communication, 2003).

[20] At each latitude band and height tested, the model receives a point for having its most probable value within 25% of the observed value for the entire year, 0.5 point for being about 25% from the mean, and nothing for being more than 25% from the mean. In general, GMI_{GCM} mixing ratios are much lower than CLAES at 1200 K, but show better agreement at 800 K, especially in the NH. The GMI_{DAS} most probable values are usually quite close to the CLAES values and the total score for the GMI_{DAS} was much higher than the GMI_{GCM}.

[21] The second and third diagnostic quantities are the phase and amplitude of the annual cycle. These quantities reflect seasonal variations in the radiative forcing and wave driving. To evaluate these aspects of the circulation independently of the mean state, model results are first scaled by the ratio of the observed/model most probable annual values. After scaling, the amplitude is judged by whether the model follows the observed annual cycle by staying within 25% of the observed values for the year. For this it receives a full point; it receives 0.5 point for deviating

Table 1. Residual Circulation Test Results

Height	Antarctic (72°–80°S)	GMI_{GCM}	GMI_{DAS}
800 K	CH_4 most probable value	0.5	0.5
800 K	Phase (1 pt) and amplitude (1 pt)	2.0	1.5
800 K	Variability	0.75	0.5
1200 K	CH_4 most probable value	0.0	0.5
1200 K	Phase (1 pt) and amplitude (1 pt)	2.0	1.5
1200 K	Variability	0.75	0.5
	Average	75%	63%
Height	Midlatitudes (35°–56°S)	GMI_{GCM}	GMI_{DAS}
380–500 K	Descent/horizontal mixing	0.67	0.33
800 K	CH_4 most probable value	0.5	1.0
800 K	Phase (1 pt) and amplitude (1 pt)	2.0	2.0
800 K	Variability	1.0	1.0
1200 K	CH_4 most probable value	0.0	1.0
1200 K	Phase (1 pt) and amplitude (1 pt)	1.0	2.0
1200 K	Variability	1.0	0.75
	Average	69%	90%
Height	Tropics (10°S to 20°N)	GMI_{GCM}	GMI_{DAS}
380–480 K	Ascent/horizontal mixing	0.80	0.85
600 K	Isolation	0.83	0.83
800 K	Isolation	1.0	0.67
1000 K	Isolation	0.83	0.50
1200 K	Isolation	1.0	0.17
	Average	89%	60%
Height	Midlatitudes (35°–56°N)	GMI_{GCM}	GMI_{DAS}
380–500 K	Descent/horizontal mixing	0.83	0.73
800 K	CH_4 most probable value	1.0	1.0
800 K	Phase (1 pt) and amplitude (1 pt)	2.0	2.0
800 K	Variability	0.85	0.75
1200 K	CH_4 most probable value	0.5	1.0
1200 K	Phase (1 pt) and amplitude (1 pt)	2.0	2.0
1200 K	Variability	1.0	0.25
	Average	91%	86%
Height	Arctic (72°–80°N)	GMI_{GCM}	GMI_{DAS}
800 K	CH_4 most probable value	1.0	0.5
800 K	Phase (1 pt) and amplitude (1 pt)	2.0	1.5
800 K	Variability	0.85	0.5
1200 K	CH_4 most probable value	0.5	1.0
1200 K	Phase (1 pt) and amplitude (1 pt)	2.0	1.5
1200 K	Variability	0.75	0.3
	Average	89%	66%

~25% from the observed cycle, and 0 for deviations greater than 25% of observed cycle. The phase is judged by requiring the model to have a minimum and maximum within a month of the observed extrema (1 point), or, when appropriate, for correctly lacking a cycle. The model receives a half point for getting only one extrema right, and nothing for a phase that bears no resemblance the observations. Both simulations did extremely well. Although the phase and amplitude are separate tests, their scores are combined in a single line in Table 1.

[22] Variability and its seasonal cycle are the fourth diagnostic quantity. Long-range transport increases CH_4 variability while mixing and photochemistry decrease it. Variability has its own seasonal cycle independent of the seasonal cycle of the most probable value. For example, the midlatitude panels in Figure 3 show almost no variation in the most probable CLAES CH_4 value, yet the breadth of the distribution varies greatly between summer and winter. In the midlatitudes in both hemispheres, the pattern of CLAES CH_4 variability shows a minimum in summer and a broad maximum in fall and winter. This indicates wave activity is strongest in fall and winter and weakest in summer. In the Antarctic, the pattern is similar but with smaller variability (weaker wave activity) near the vortex, especially at 800 K in winter (not shown). In the Arctic, variability is greatest in winter and quite low in summer when photochemistry is faster.

[23] This semi-quantitative test looks at the breadth of the model distribution in each season and looks for agreement with the seasonal cycle of variability. The model values are scaled as before, so that variability is judged independently of the annual mean. The GMI_{GCM} overall shows a very good cycle of CH_4 variability at all latitudes and heights tested. It produces a summer minimum in variability (i.e., there is no excess in wave-driven transport that interferes with the reduction of variability by fast photochemistry) and a maximum in variability in the appropriate cool seasons. Its grades are less than perfect in the polar regions because winter wave driving appears to be a little weak. This is consistent with the results of the most probable value test, where low CH_4 at 1200 K also pointed to weak wave driving. The GMI_{GCM} midlatitude cycle of variability is in very good agreement with CLAES.

[24] The GMI_{DAS} does not consistently show the right cycle of CH_4 variability. Like the GMI_{GCM}, wave driving in the polar regions is weak, but it receives lower grades because of noisiness in summer, especially in the Northern Hemisphere. Like the $GMI_{GCM,}$ the GMI_{DAS} does best in the midlatitudes, with the exception of 1200 K in the north. There it has a minimum of variability in late winter and very large variability in summer, looking nothing like the observations. The large summer variability seen here could be caused by excessive poleward transport from low latitudes. This is also examined by the test in section 2.2.2.3. The grading for this test is similar to previous tests: A full point is awarded for deviations less than 25% from the observations while deviations greater than 25% result in no points.

2.2.2.2. Seasonal Variations in Lower Stratospheric N_2O

[25] Lower stratospheric profiles of the long-lived tracer N_2O reflect a balance between the diabatic circulation and meridional mixing. N_2O measurements in the midlatitude and tropics, derived from 11 years of ER-2 N_2O data, were used to create seasonal mean profiles for this test [*Strahan et al.*, 1999]. This test and the details of the grading are described by *Douglass et al.* [1999]. Figure 4 shows examples of the model/data agreement and Table 1 gives the models' scores. Both simulations perform acceptably in the tropics and northern midlatitudes, and neither has excellent agreement in the Southern Hemisphere (SH). The GMI_{GCM} is closer to observations than the GMI_{DAS}.

2.2.2.3. Tropical Isolation in the Middle and Upper Stratosphere

[26] This is a variation of "test 3" of *Douglass et al.* [1999], in which the bimodality of CLAES N_2O pdfs between 10°S and 45°N were used to assess a model's ability to produce a sufficient tropical/midlatitude barrier. The barrier results from strong shear in the subtropics that confines wave breaking to the midlatitudes (the "surf zone") [*Polvani et al.*, 1995]. Ascending air of recent tropospheric origin results in high tracer values on the tropical side of the barrier, while descent of older air on the poleward side gives lower tracer mixing ratios in the surf zone. In the middle and upper stratosphere, the phase of the QBO as well as seasonal variations in wave driving

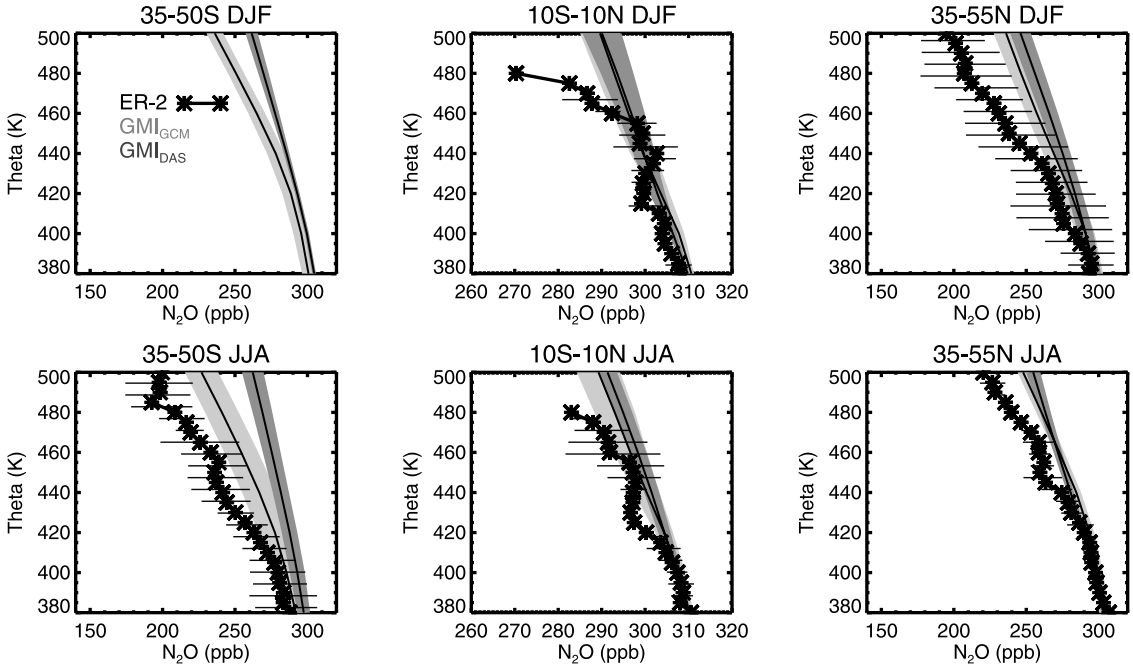

Figure 4. Comparison of modeled and observationally derived seasonal mean N_2O profiles in the lower stratosphere. The panels show comparisons for three latitude ranges and two seasons. The shading around the solid lines represents 1σ of the models' variability for that latitude and season; 1σ of the observational variability is shown by the horizontal bars. Both models are consistently higher than the observations at and above 420 K, with the GMI_{GCM} profiles usually lying closer to the observations than the GMI_{DAS}.

cause variations in the barrier strength and location, affecting the distinctiveness of tropical and midlatitude air masses [*Gray and Russell*, 1999]. The original test was conducted on 3 pressure surfaces from 31–7 hPa; here we use 4 theta surfaces from 600 to 1200 K. A wide latitude range is chosen so that the QBO phase-dependent location of the subtropical barrier will not affect the modality of the pdf. The phase-dependent secondary circulation set up by the QBO also causes significant interannual variability in constituent mixing ratios [*Gray and Russell*, 1999]; thus we compare only the modality of the distribution and not the mixing ratios or the absolute separation of the peaks.

[27] The models are graded every 200 K from 600 K to 1200 K. The CLAES N_2O pdfs show isolation of the tropics at all levels and in all seasons compared. (Spring is excluded because the observed N_2O distribution was non-

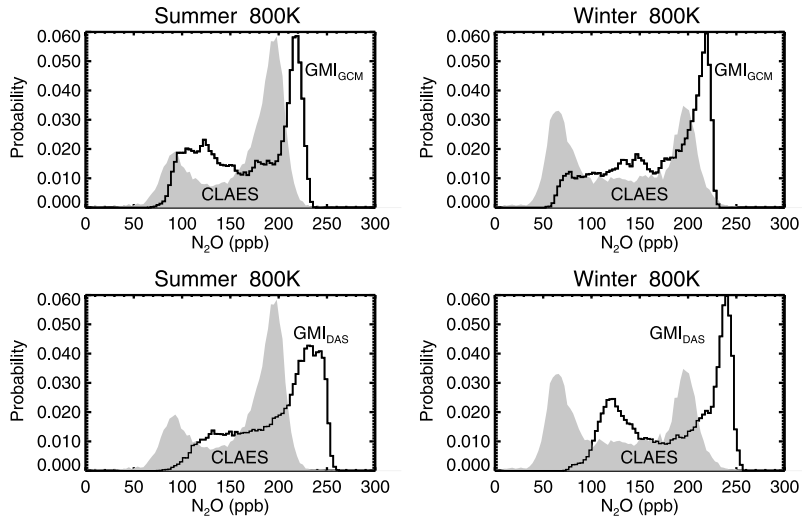

Figure 5. Separation of tropical and midlatitude air masses shown by CLAES N_2O pdfs, and model comparisons on the 800 K surface in summer and winter. While both models maintain separation in winter at all levels examined (i.e., a bimodal distribution), only the GMI_{GCM} consistently keeps a clear separation in summer and fall at all levels.

Table 2. Summary of Transport Performance[a]

	Comparison of Simulations by Height and Hemisphere			
	SH, GMI_{GCM}	SH, GMI_{DAS}	NH, GMI_{GCM}	NH, GMI_{DAS}
Upper stratosphere	59%	78%	84%	76%
Middle stratosphere	84%	81%	96%	78%
Lower stratosphere	67%	33%	83%	73%
	Quality of Mixing Barriers			
	GMI_{GCM}	GMI_{DAS}		
Antarctic vortex	fair	poor		
Arctic vortex	fair	poor		
Tropics (600–800 K)	very good	fair		
Tropics (1000–1200 K)	very good	poor		
Tropopause: 60°N	very good	very good		

[a]NH, Northern Hemisphere; SH, Southern Hemisphere.

stationary.) A simulation is granted 1 point for producing two peaks separated by a minimum, even a weak one. A half point is given for a tropical peak with a long midlatitude tail instead of a clear minimum, and no point is given if a single, short-tailed (i.e., well mixed) peak is found. Figure 5 provides examples of the performance of these simulations at 800 K. The GMI_{GCM} maintains a tropical/midlatitude separation in all 3 seasons (scoring 92%), while the GMI_{DAS} makes a clear separation only in winter (scoring 54%). The GMI_{GCM} performance did not depend on height, while the GMI_{DAS} showed decreasing tropical isolation with increasing height (see Table 1).

2.2.3. Upper Troposphere/Lower Stratosphere Separation

[28] This simple test is important because it gauges whether a model has the correct pathway of transport from the upper troposphere to lower stratosphere. Referred to as "test 5" by *Douglass et al.* [1999], it examines the phase lag of the CO_2 seasonal cycle across the tropopause at 60°N. Models must show at least a 2-month lag between the CO_2 seasonal cycle maximum on the highest tropospheric and the lowest stratospheric levels. The presence of the lag indicates that air from the extratropical upper troposphere does not go directly up into the lower stratosphere, but takes a path to the stratosphere via the tropical tropopause [*Boering et al.*, 1996; *Strahan et al.*, 1998]. This test is useful for identifying simulations with excessive convective transport. In this study, both simulations pass (Table 2).

2.3. Specialized Tests Relevant to Ozone Simulations

2.3.1. Spatial and Temporal Coverage of PSC-Producing Temperatures

[29] In regions where heterogeneous chemical reactions occur, the distribution of temperatures is more important than the mean. *Austin et al.* [2003] evaluated model polar temperatures by calculating the product of models' areal and temporal coverage of NAT- and ice-forming temperatures in each hemisphere and comparing that with the quantity derived from observations. Similarly, in this test we judge a model by whether it can produce a spatially realistic distribution of NAT-forming temperatures during the appropriate months. Since polar ozone loss depends on both low temperatures and sunlight, this test is designed to look specifically at the latitudes and months where PSC-forming temperatures are reached. The score for this test is a description of the model's behavior compared to climatology, providing context for the interpretation of a model's polar processes.

[30] This test uses 20 years of daily NCEP/NCAR reanalysis to first calculate a climatological temperature distribution for 3 latitude bands (70°–90°, 60°–70°, and 50°–60°) at 6 pressure levels from 150 to 10 hPa in each hemisphere. On the basis of the PSC frost point temperature for each pressure level (assuming 4 ppmv H_2O), we then calculate the fractional area of each latitude band that is covered by temperatures at or below the frost point. Using all 20 years of data, the monthly mean and standard deviation of this fractional area are calculated. Comparisons between the modeled and observed fractional areas are made for several one-month periods that span the cold season (May–November in the Antarctic, December–March in the Arctic). Since low-temperature bias in the Antarctic stratosphere is a longstanding model issue [*Pawson et al.*, 2000], this test serves to assess whether a model could produce PSCs outside of the expected latitudes, heights, and months.

[31] Rather than assigning a numerical score to the simulations, we use this test to label a model's potential to form PSCs. Model fractional coverage within 1.5 standard deviations (σ) of the climatological fraction is labeled "normal"; fractional coverage between 1.5σ and 3σ above (below) the observed mean is considered "colder (warmer) than normal," and 3σ above (below) the mean is "much colder (warmer) than normal" and possibly unphysical. It's important to remember that this test is not strictly a temperature test. For example, a model with realistic July temperatures in the Antarctic would produce the same fractional area of PSC-producing temperatures as an unrealistically cold simulation (1.0), both in agreement with observations. At the same time, a model that produces fractional coverage 3σ greater than observations in May may have only a few percent of its area covered by PSCs.

[32] Figure 6 characterizes the PSC-producing potential in the FVGCM and FVDAS from May to November. In fall, winter, and spring, both models produce climatologically normal coverage of PSC-producing temperatures from 50° to 90°S and 150 to 10 hPa most of the time, but usually some part of the polar region is colder than normal each month. The FVGCM tends to be a little cold at 100 hPa and below, especially at latitudes near the edge and outside the vortex. Both models are colder than normal inside the vortex near 30 hPa in late winter, which could lead to greater ozone loss there. The information

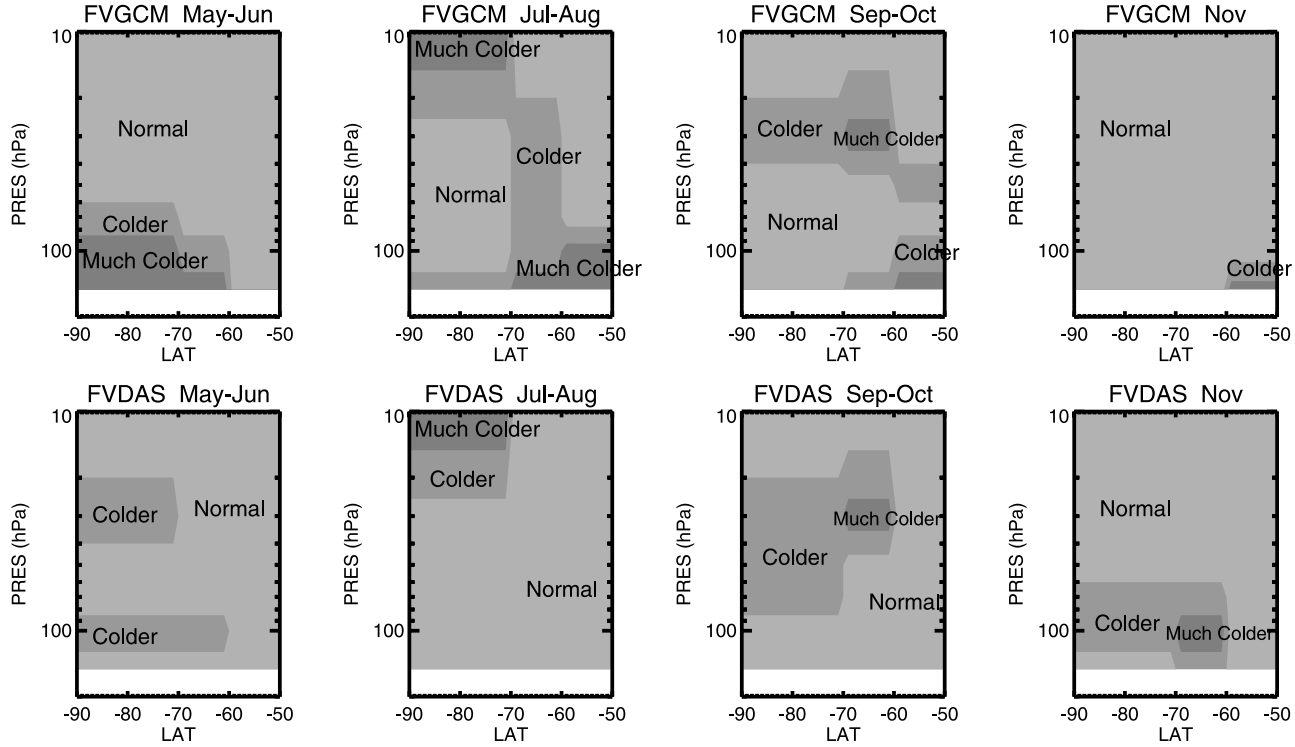

Figure 6. Distribution of normal, below normal (1.5–3σ), and much below normal (>3σ) model temperatures in the Antarctic from 50° to 90°S, 150 to 10 hPa. Both models have large areas of climatologically normal temperatures. The panels span the entire cold season for the Antarctic, and each comparison spans a one-month interval.

presented in this figure provides a context for the interpretation of ozone behavior given by Considine et al. (submitted manuscript, 2004) and Douglass et al. (submitted manuscript, 2004).

[33] The same analysis was performed for December through March in the Northern Hemisphere. The Arctic has higher mean temperatures and much larger interannual temperature variability than the Antarctic, such that being within 1.5 of the mean can mean having some PSCs or none at all. Figure 7 characterizes the Arctic coverage of PSC-producing temperature in the models. Much of the FVGCM vortex and edge region can be labeled as normal. Temperatures are on the warm side of the mean in early winter (no PSCs) while it is quite cold from 70 to 150 hPa in late winter. However, the absolute effect on PSC areal coverage is small: For example, ''much colder than normal'' at 100 hPa in March still means less than 10% coverage compared to about 2% climatologically. The FVDAS temperatures are climatologically normal much of the time, with the exception that February and March are much colder than average at lower levels. The FVDAS assimilated wind fields represent the period 1 July 1999 to 30 June 2000, reflecting the unusually low stratospheric temperatures observed in March 2000 [*Newman et al.*, 2002].

[34] The particular FVGCM year evaluated here, part of a 35-year FVGCM integration, was selected because it had a cold Arctic winter with conditions similar to those observed in the winter of 1999–2000, the year of the assimilated data set. It is interesting to note that this test rates the FVGCM winter as climatologically normal. An average FVGCM winter from this multi-year integration would probably be rated warmer than normal.

2.3.2. Lower Stratospheric Vortex Behavior During Breakdown

[35] The huge ozone losses observed every spring in the Antarctic are the result of a unique combination of dynamics and chemistry found nowhere else. A model's ability to realistically simulate vortex erosion and mixing processes during breakdown may be crucial to its credibility in predicting how declining halogens will affect the depth of the ozone hole and the dispersion of ozone-depleted air. For example, if a model brings midlatitude air into the vortex in early spring, this intrusion of nondenitrified air would cause Cl-catalyzed loss processes to shut down prematurely. Ozone would not get realistically low inside the vortex, and vortex air dispersed to lower latitudes would have more ozone. This test gauges vortex isolation and exchange between the vortex and midlatitudes in spring.

[36] Methane measurements from HALOE and CLAES are both suitable for developing a mixing and isolation diagnostic, but the HALOE data provide a more complete picture. CLAES has good spatial coverage down to 80°S for a month at a time, but there are no measurements in October and only 1 austral spring was sampled. CLAES CH_4 data also have large uncertainty at 450 K where ozone loss rates are greatest [*Roche et al.*, 1996; Considine et al., submitted manuscript, 2004]. While HALOE obtains only 15 profiles a day in a hemisphere, it has done so for a decade; low interannual variability in the Southern Hemisphere allows those measurements to be sensibly combined to derive a

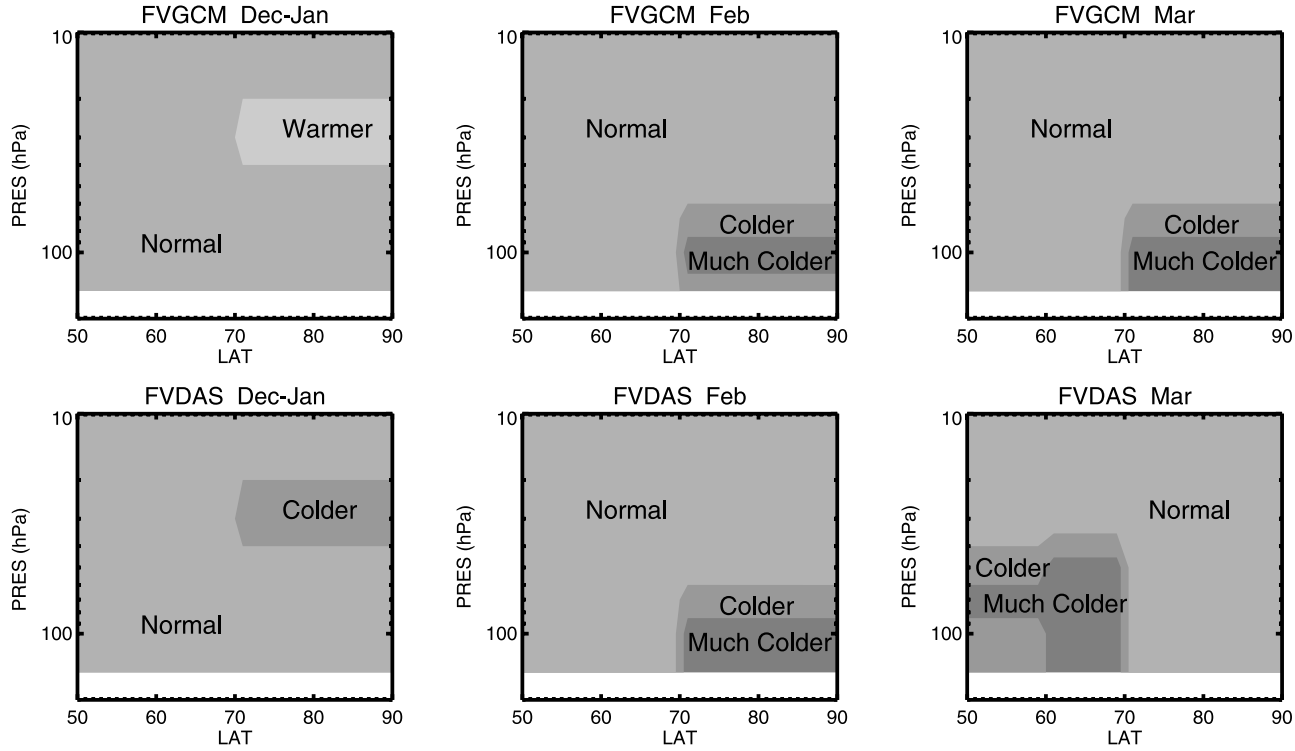

Figure 7. Same as Figure 6, except for the Arctic winter. The FVDAS is unusually cold in March, reflecting Arctic conditions in 2000. Overall, the FVGCM is warmer in the Arctic and makes fewer PSCs than does the FVDAS, but because variability in the Arctic is so great, both models are categorized as climatologically normal.

mean dynamical picture of vortex development and breakdown for the entire austral spring. (HALOE data from 2002, a dynamically unusual year, are not included in this test.) The two CH_4 data sets can be compared on the 600 K surface in September and November. In a prototype test, HALOE and CLAES CH_4 pdfs exhibited the same dynamical features, lending confidence to the use of HALOE data, which has far less spatial coverage in any single year.

[37] The test examines the springtime evolution of the CH_4 pdfs of two latitude bands, 60°–80°S and 40°–60°S, on the 450 K and 600 K surfaces. Pdfs are derived by binning 8 years of measurements from each latitude range for each month of spring. The 60°–80°S range is almost strictly vortex air in early spring, with a small but isolated vortex core lasting into November. The 40°–60°S band is almost strictly midlatitude, or "surf zone," air. The dynamics of vortex breakdown and the extent and direction of mixing between the vortex and surf zone are revealed by several features of the pdfs: the separation of the peaks, the depth of the minimum between the peaks, and changes in the mean and most probable values of the peaks during spring.

[38] The evolution of HALOE CH_4 pdfs, shown in the middle column of Figure 8, reveals the process of vortex breakdown in the Antarctic lower stratosphere. The September data show two broad but distinct distributions with peaks separated by about 400 ppb; the lack of a deep minimum indicates a strong barrier to mixing has not yet developed. By October and November a deep minimum has developed, indicating a barrier to mixing has formed. Mixing within the vortex also increases during these months as indicated by the narrower vortex distribution. The most probable value decreases, but because CH_4 decreases toward the pole, this may reflect the higher mean sampling latitude in October (69°S) compared to 64°S in September (67°S in November). The most probable value in the vortex declines more than 30 ppb between October and November, arguing strongly against any intrusion of surf zone air, which is typically 500 ppb higher; even a narrow band of mixing at the vortex edge would result in an increase of the vortex mean. Diabatic heating is near zero in spring [*Rosenfield et al.*, 1994; *Rosenfield and Schoeberl*, 2001], and thus descent is not expected to be important. The bimodal distribution of the November 60°–80°S data indicates the presence of both vortex and nonvortex air, where in September there was only vortex air, illustrating the shrinking area of the vortex. The high CH_4 peak of the 60°–80°S November distribution is nearly identical to the 40°–60°S peak, suggesting that air exiting the vortex becomes rapidly mixed into (i.e., has the same characteristics as) the surf zone. The development of this bimodal structure in the 60°–80°S band and the endurance of the low CH_4 (vortex) peak indicate that (1) the vortex persists throughout spring, (2) it breaks down by erosion, and (3) it maintains its identity during breakdown, with no evidence of midlatitude air mixing in. The HALOE data at 600 K, not shown, give a similar picture of breakdown and show an even stronger barrier to mixing.

[39] This data set provides a clear picture of vortex breakdown and an excellent basis for model evaluation of

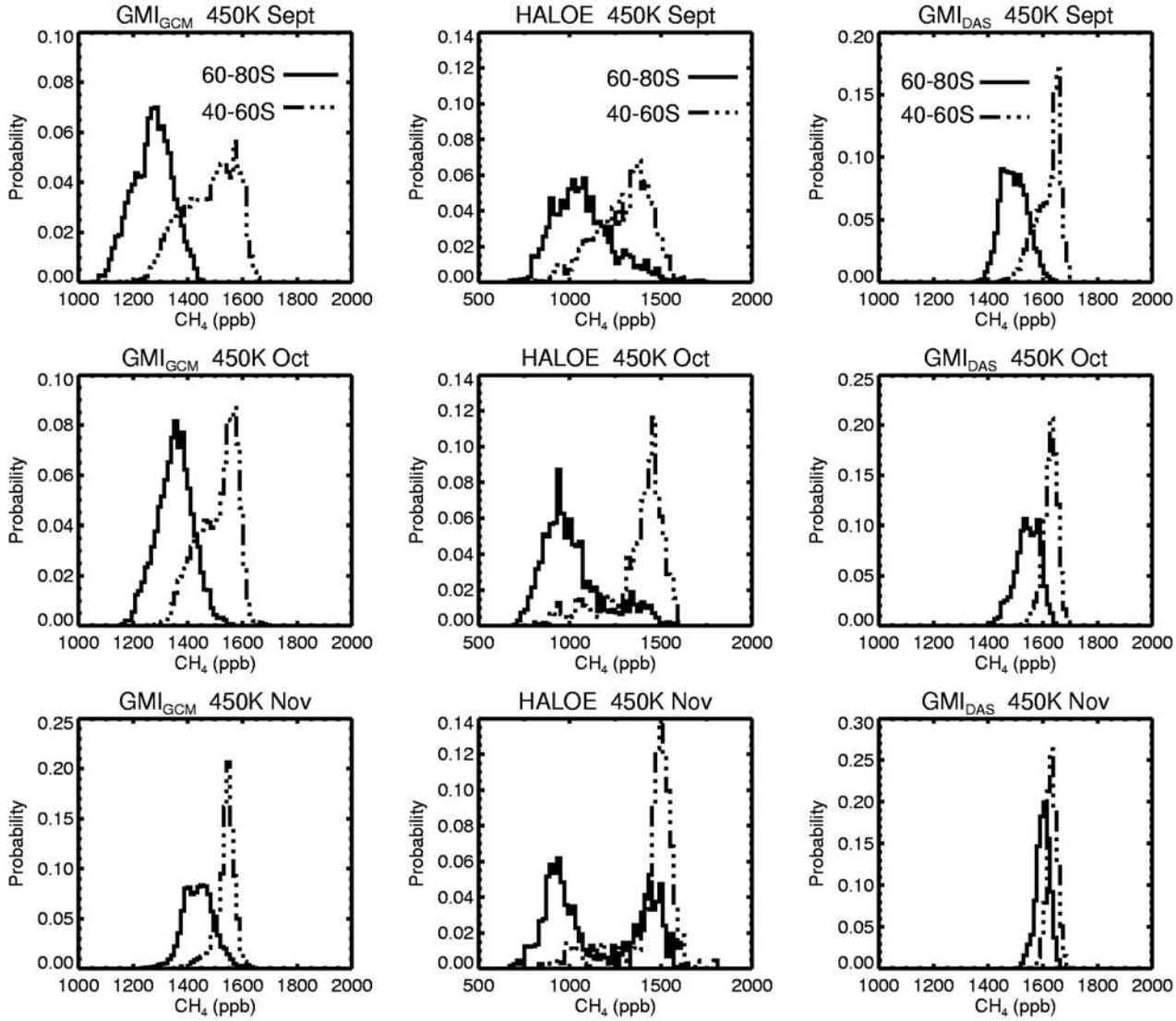

Figure 8. Evolution of CH_4 distributions on the 450 K surface inside and outside the Antarctic vortex in spring. The central column, representing an 8-year accumulation of HALOE observations in austral spring, demonstrates that the vortex air mass maintains its identity while gradually eroding. The GMI_{GCM} simulation (left column) maintains some separation through the spring, but large overlap between the distributions indicates exchange between the regions, in contrast to the observations. The GMI_{DAS} simulation (right column) does a worse job of maintaining separation, and by November the vortex and midlatitudes are nearly identical (i.e., well mixed).

this process. The left column of Figure 8 shows the breakdown of the GMI_{GCM} vortex. The evolution of the model breakdown differs in many ways from the observations. While the September distributions have an acceptable 400 ppb separation, that separation decreases in the following months, both distributions narrow (no long tails exist at any time), no deep and wide minimum forms between the peaks, and the most probable vortex value increases by nearly 200 ppb in stark contrast to the 100 ppb decrease seen in the observations. Vortex evolution in the GMI_{DAS} simulation is very similar. The separation of the GMI_{DAS} distributions is smaller to begin with and shrinks rapidly in spring. By November, the GMI_{DAS} vortex is not distinct from the midlatitudes.

[40] In both GMI-CTM simulations, the decreasing separation of the peaks and the lack of a deep minimum between them indicate that a strong barrier to mixing never forms. The separation of the peaks of the HALOE vortex and midlatitude distributions increases during spring at 450 K and 600 K, while the peaks of the GMI_{GCM} and GMI_{DAS} vortex and midlatitude distributions move toward each other. The HALOE most probable midlatitude CH_4 increases from 1370 to 1500 ppb because of wave-driven mixing in the surf zone, which brings higher CH_4 poleward from low latitudes. At the same time, the continued isolation of the vortex is demonstrated by the decrease in $60°-80°S$ CH_4. In contrast to observed behavior, the models' midlatitude peak moves lower and

the vortex peak moves higher, suggesting a continuous exchange of air between 40°–60°S and 60°–80°S as the vortex shrinks; the GMI_{DAS} has stronger exchange than the GMI_{GCM} because its distributions are merged by November. The HALOE pdfs clearly show that the vortex breaks down by erosion without entrainment of midlatitude air, revealing a fundamental difference between modeled and observed vortex behavior. By November, the GMI_{DAS} CH_4 distributions indicate a nearly homogeneous region from 40° to 80°S, in great contrast to the HALOE pdfs that indicate the persistence of a small, well-isolated vortex.

[41] The large ozone losses observed each October in the Antarctic are possible only because of the strict isolation of chemically perturbed air inside the vortex. This unique requirement of both chemical processing and isolation cannot be met in a model that lacks a strong barrier to mixing; the lack of vortex isolation in the simulations compromises their ability to sequester chemically perturbed air. Poor performance on this test suggests significant consequences for model vortex mixing ratios of Cl_y, ClO, and NO_x, and for ozone loss. If the model vortex permits entry of NO_x-rich/Cl_y-poor air from the midlatitudes before breakdown, it will not be able to achieve near total destruction of O_3 in the lower stratospheric Antarctic vortex. Model ozone loss will be incomplete, and the excessive exchange between the vortex and midlatitudes will disperse that air prematurely to the Southern Hemisphere.

[42] In contrast to the GMI_{GCM} results presented here, the Antarctic vortex was well isolated in two experiments using FVGCM meteorological fields from a different year. One experiment used the Goddard CTM while the other carried CH_4 online in the FV general circulation model; both experiments used the same year of FVGCM winds at their native 2° × 2.5° horizontal resolution. Low interannual variability in the FVGCM Southern Hemisphere suggests that the choice of a different meteorological year is not the cause of the differences with the GMI_{GCM}. We are currently investigating the details of these simulations to diagnose the cause for these large differences in vortex permeability. Differences between these simulations include the horizontal and temporal resolution of the winds and limits imposed on the vertical fluxes in the CTM.

[43] Figure 9 compares model and HALOE Arctic pdfs on the 600 K surface. There are no HALOE high-latitude measurements in February. Although the Arctic vortex in March is much smaller than the Antarctic in September, the observations still show a clear separation between the tiny vortex and the midlatitudes. The GMI_{GCM} manages to keep some separation between the air masses, but the GMI_{DAS} shows a completely homogenized region from 40° to 80°S by March; the observations indicate mixing is still incomplete in April. The GMI_{GCM} has broad distributions in February and March while the GMI_{DAS} distributions are much narrower, indicating stronger horizontal mixing. In a typical Arctic winter with small O_3 losses, excessive mixing across the vortex edge will have little impact on the O_3 distribution since its horizontal gradients are fairly flat in the 500–600 K range [*Strahan*, 2002]. However, should the Arctic stratosphere have a cooling trend in the 21st century with concomitantly larger wintertime O_3 losses, the lack of vortex isolation such as shown here may invalidate model predictions.

3. Grading Summary: Model Credibility

[44] Most air enters the stratosphere through the tropical tropopause. Both simulations begin this journey reasonably well, with good agreement between modeled and observed N_2O profiles in the tropical lower stratosphere. At 600 K and above, the GMI_{DAS} is unable to maintain an isolated tropical air mass in summer and fall, unlike the GMI_{GCM}, which maintains a distinct tropical air mass at all altitudes and seasons tested. The weak tropical barrier in the GMI_{DAS} allows too much N_2O into the midlatitudes, producing poor comparisons with lower stratospheric N_2O profiles (e.g., Figure 4). *Douglass et al.* [2003] have also diagnosed excessive mixing between the tropics and midlatitudes in a simulation using FVDAS winds in the Goddard CTM. They argue that the excessive horizontal mixing probably results from the insertion of observed winds and temperatures in the assimilation system, which imposes an additional forcing on the equations of motion. The GMI_{DAS} tropical isolation gets worse with height and leads to problems with midlatitude and high-latitude tracer distributions and variability.

[45] The test of CH_4 annual cycles in the southern extratropics shows insufficient wave driving in the austral fall. Wave driving brings high CH_4 from the low latitudes to the polar region. The inadequacy of the wave-driven transport is seen in the GMI_{GCM} Antarctic upper stratosphere in the form of too little CH_4, low variability, and almost no CH_4 increase in fall and winter compared to CLAES (Figure 3, 1200 K, 72°–80°S). Similar CH_4 behavior is seen in an FVGCM simulation with online chemistry, suggesting that the processes controlling CH_4 at these levels do not strongly depend on CTM details such as the horizontal resolution and model lid height. The GMI_{DAS} shows more of this transport occurring. When the GMI_{GCM} vortex forms in the fall and descent begins in the upper stratosphere, CH_4 values lower than observed are trapped in the descending vortex. By late winter, CH_4 in the GMI_{GCM} Antarctic vortex is in close agreement with CLAES (600–1200 K), indicating some midlatitude air has mixed into the vortex during descent; by spring, model CH_4 has become too high.

[46] In Figure 8, which shows the separation of vortex and midlatitude air, the vortex most probable value rises in the model while decreasing in the observations. The GMI_{DAS} shows the same feature but starts with even higher CH_4 in the vortex. As previously discussed, these test results indicate that the lack of isolation allows too much high-CH_4 air from middle latitudes into the vortex. However, because an online FVGCM experiment has demonstrated the model's ability to produce nearly realistic vortex isolation year after year, we suspect that the lack of vortex isolation in the GMI simulations implicates the CTM horizontal or temporal resolution.

[47] The overall result of the middle- and high-latitude residual circulation tests is that tracer transport in the offline GMI_{GCM} has greater fidelity throughout the stratosphere than the GMI_{DAS}. The GMI_{GCM} has greater realism in its Northern Hemisphere than its Southern Hemisphere and it performs best in the middle stratosphere. The Southern

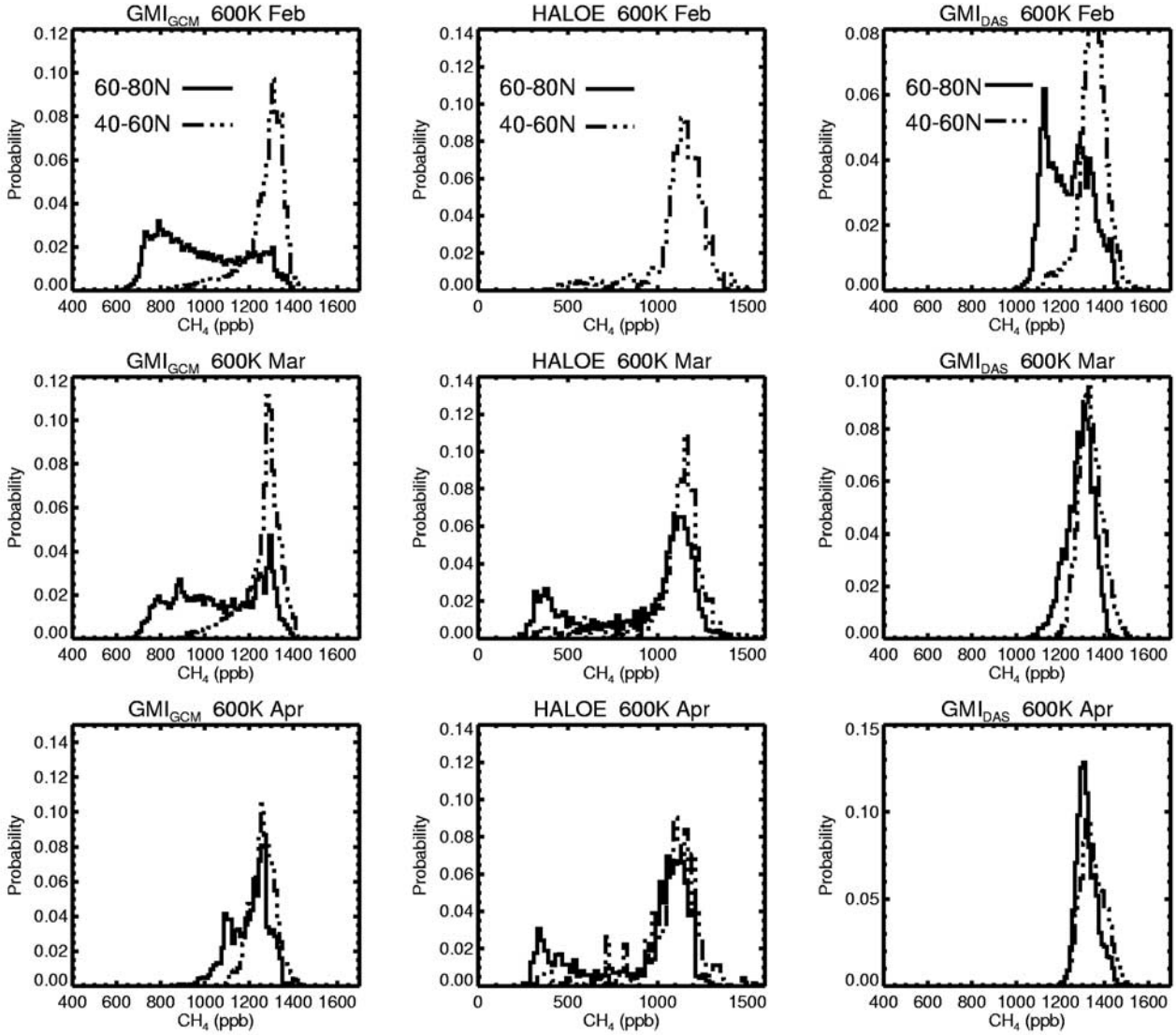

Figure 9. Evolution of CH_4 distributions on the 600 K surface inside and outside the Arctic vortex in spring. The GMI_{GCM} distributions (left column) show good separation in February and March, with near total mixing by April. HALOE data, accumulated over 8 years (center column), show a small, distinct vortex in April; no high-latitude data are available in February. The slight separation between regions shown in February in the GMI_{DAS} simulation (right column) is completely erased by March. A substantial vortex existed well into March 2000 [*Newman et al.*, 2002], but the GMI_{DAS} vortex is not distinct from the midlatitudes in this simulation.

Hemisphere upper stratosphere is where the GMI_{GCM} has the worst comparison with observations and where the GMI_{DAS} scores considerably higher. The GMI_{DAS} performs best in the midlatitudes, north and south, but struggles with the Southern Hemisphere lower stratosphere. In the polar lower stratosphere, neither simulation is able develop an impermeable vortex edge. Table 2 summarizes residual circulation grading as a function of height and hemisphere, and rates barrier formation ability.

[48] Temperature-dependent gas phase reactions are likely to be carried out at the right rates in both simulations. At 30 hPa and below, each model achieves climatologically normal temperatures at nearly all latitudes and seasons. Higher up, both models are biased slightly low, but near the stratopause both models are a few degrees too high. Overall, the FVDAS is always closer to NCEP 20-year climatological temperatures than the FVGCM in the stratosphere.

[49] Both simulations do a good job of producing realistic areal coverage of PSC-forming temperatures as a function of month, latitude, and altitude. The region of FVGCM PSC-producing temperatures extends a little too far from the pole, particularly at 100 hPa and below. The FVDAS Antarctic lower stratosphere warms a little later in spring than the climatological average. In the Arctic, the FVGCM was climatologically near normal at 50 hPa and above, but cold in the late winter vortex at 70 hPa and below. The FVDAS was near normal in much of the Arctic, but much below normal in March outside the vortex near 70 hPa, reflecting the cold stratospheric spring of 2000.

[50] Overall, the GMI_{GCM} does a better job of barrier formation, particularly in the tropics though only moderately so in the polar regions. This is consistent with the model mean age of air comparison showing the GMI_{GCM} to have older air in the polar lower stratosphere than the GMI_{DAS} (Considine et al., submitted manuscript, 2004). The primary weakness in the GMI_{GCM} appears in the Southern Hemisphere spring and fall, the seasons when the greatest wave activity should occur [*Randel*, 1988]. The GMI_{DAS} does a better job there, suggesting that the insertion of observations in the DAS may improve GCM deficiencies in this region.

4. Final Thoughts on Model Evaluation

[51] No model can faithfully represent all known atmospheric processes, but by understanding both the skills and the deficiencies of a model, one can determine its best use. Transport and chemistry influence the distribution of ozone at all altitudes and latitudes of the stratosphere, requiring a model to perform well just about everywhere. To study the effect of changing chlorine levels on stratospheric ozone, a model requires additional testing in regions where chlorine plays a significant role in ozone loss (i.e., the upper stratosphere and the polar lower stratosphere). This reflects the philosophy of evaluation used here.

[52] These evaluations provide insight into the usefulness of offline chemistry and transport simulations using the FVGCM and FVDAS meteorological fields. The quality of these simulations is affected by all aspects of the model implementation, including the input meteorological fields, the offline advection scheme, the resolution of the CTM, and the choice of the chemical mechanism and solver. Differences in the time step of the chemical mechanism and the advection scheme will lead to interactions between these modules, especially for diurnally varying species near the terminator. This leads to inherent differences in performance between offline and online chemistry. Experiments performed at $2^\circ \times 2.5^\circ$ horizontal resolution will not give the same results as a $4^\circ \times 5^\circ$ experiment, especially for meridional tracer gradients and for the CH_4 vortex mixing diagnostic. Experiments with online parameterized CH_4 chemistry in the FVGCM revealed that tracer transport is less diffusive and more realistic online than with the same meteorological fields in the offline model. Using the model diagnostics shown here, sensitivity of results to resolution and implementation choices can be tested, allowing the user to select simulations with the greatest fidelity to physical processes. Objective evaluation of model processes, using diagnostics such as those presented here and by *Douglass et al.* [1999], provides a way to reduce uncertainty in model calculations.

[53] **Acknowledgments.** We thank David Considine for insightful comments and suggestions. This work was supported by the NASA Atmospheric Chemistry, Modeling, and Analysis Program.

References

Austin, J., et al. (2003), Uncertainties and assessments of chemistry-climate models of the stratosphere, *Atmos. Chem Phys.*, *3*, 1–27.

Boering, K. A., S. C. Wofsy, B. C. Daube, J. R. Schneider, M. Loewenstein, J. R. Podolske, and T. J. Conway (1996), Stratospheric mean ages and transport rates from observations of CO_2 and N_2O, *Science*, *274*, 1340–1343.

Considine, D. B., A. R. Douglass, P. S. Connell, D. E. Kinnison, and D. A. Rotman (2000), A polar stratospheric cloud parameterization for the global modeling initiative three-dimensional model and its response to stratospheric aircraft, *J. Geophys. Res.*, *105*, 3966–3973.

Danilin, M. Y., et al. (1998), Aviation fuel tracer simulation: Model intercomparison and implications, *Geophys. Res. Lett.*, *25*, 3947–3950.

Douglass, A. R., M. J. Prather, T. M. Hall, S. E. Strahan, P. J. Rasch, L. C. Sparling, L. Coy, and J. M. Rodriguez (1999), Choosing meteorological input for the global modeling initiative assessment of high-speed aircraft, *J. Geophys. Res.*, *104*, 27,545–27,564.

Douglass, A. R., M. R. Schoeberl, R. B. Rood, and S. Pawson (2003), Evaluation of transport in the lower tropical stratosphere in a global chemistry and transport model, *J. Geophys. Res.*, *108*(D9), 4259, doi:10.1029/2002JD002696.

Gray, L. J., and J. M. Russell (1999), Interannual variability of trace gases in the subtropical winter stratosphere, *J. Atmos. Sci.*, *56*, 977–993.

Kinnison, D. E., et al. (2001), The Global Modeling Initiative assessment model: Application to high-speed civil transport perturbation, *J. Geophys. Res.*, *106*, 1692–1712.

Lin, S. J., and R. B. Rood (1996), Multidimensional flux form semi-Lagrangian transport schemes, *Mon. Weather Rev.*, *124*, 2046–2070.

Newman, P. A., E. R. Nash, and J. E. Rosenfield (2001), What controls the temperature of the Arctic stratosphere during the spring?, *J. Geophys. Res.*, *106*, 19,999–20,010.

Newman, P. A., et al. (2002), An overview of the SOLVE/THESEO 2000 campaign, *J. Geophys. Res.*, *107*(D20), 8259, doi:10.1029/2001JD001303.

Park, J. H., et al. (1996), Validation of halogen occultation experiment CH_4 measurements from the UARS, *J. Geophys. Res.*, *101*, 10,183–10,205.

Park, J. H., M. K. W. Ko, C. H. Jackman, R. A. Plumb, J. A. Kaye, and K. H. Sage (1999), Models and measurements intercomparison II, *NASA Tech. Memo., NASA/TM-1999-209554*.

Pawson, S., et al. (2000), The GCM-Reality Intercomparison Project for SPARC (GRIPS): Scientific issues and initial results, *Bull. Am. Meteorol. Soc.*, *81*, 781–796.

Polvani, L. M., D. W. Waugh, and R. A. Plumb (1995), On the subtropical edge of the stratospheric surf zone, *J. Atmos. Sci.*, *52*, 1288–1309.

Randel, W. J. (1988), Seasonal evolution of planetary waves in the Southern Hemisphere stratosphere and troposphere, *Q. J. R. Meteorol. Soc.*, *114*, 1385–1409.

Randel, W. J., F. Wu, J. M. Russell, A. Roche, and J. W. Waters (1998), Seasonal cycles and QBO variations in stratospheric CH_4 and H_2O observed in UARS HALOE data, *J. Atmos. Sci.*, *55*, 163–185.

Roche, A. E., et al. (1996), Validation of CH_4 and N_2O measurements by the CLAES instrument on the Upper Atmosphere Research Satellite, *J. Geophys. Res.*, *101*, 9679–9710.

Rosenfield, J. E., and M. R. Schoeberl (2001), On the origin of polar vortex air, *J. Geophys. Res.*, *106*, 33,485–33,497.

Rosenfield, J. E., P. A. Newman, and M. R. Schoeberl (1994), Computations of diabatic descent in the stratospheric polar vortex, *J. Geophys. Res.*, *99*, 16,677–16,689.

Rotman, D. A., et al. (2001), Global Modeling Initiative assessment model: Model description, integration, and testing of the transport shell, *J. Geophys. Res.*, *106*, 1669–1692.

Schoeberl, M. R., A. R. Douglass, Z. Zhu, and S. Pawson (2003), A comparison of the lower stratospheric age spectra derived from a general circulation model and two data assimilation systems, *J. Geophys. Res.*, *108*(D3), 4113, doi:10.1029/2002JD002652.

Sparling, L. C. (2000), Statistical perspectives on stratospheric transport, *Rev. Geophys.*, *38*, 417–436.

Strahan, S. E. (2002), Influence of planetary wave transport on Arctic ozone as observed by Polar Ozone and Aerosol Measurement (POAM) III, *J. Geophys. Res.*, *107*(D20), 4417, doi:10.1029/2002JD002189.

Strahan, S. E., A. R. Douglass, J. E. Nielsen, and K. A. Boering (1998), The CO_2 seasonal cycle as a tracer of transport, *J. Geophys. Res.*, *103*, 13,729–13,742.

Strahan, S. E., M. Loewenstein, and J. R. Podolske (1999), Climatology and small-scale structure of lower stratospheric N_2O based on in situ observations, *J. Geophys. Res.*, *104*, 2195–2208.

World Meteorological Organization (WMO) (2002), Scientific assessment of ozone depletion: 2002, *WMO Publ. 47*, Geneva.

A. R. Douglass, NASA Goddard Space Flight Center, Code 916, Greenbelt, MD 20771, USA. (anne.r.douglass@nasa.gov)

S. E. Strahan, Goddard Earth Science and Technology Center, University of Maryland Baltimore County, Baltimore, MD 21250, USA. (strahan@code916.gsfc.nasa.gov)